Access control in NB-IoT networks: a deep reinforcement learning strategy

Yassine Hadjadj-Aoul

ACCESS CONTROL IN NB-IOT NETWORKS: A DEEP REINFORCEMENT LEARNING STRATEGY

Yassine Hadjadj-Aoul
Associate professor, Univ Rennes
IRISA/INRIA Dionysos team-project

Lien pour les questions: shorturl.at/prNW5

Thursday, November 26th, 2020

PLAN

Introduction

Access overview of IoT devices

A model for the access

Efficient support of a massive number of IoT devices using reinforcement learning

Conclusions

INTRODUCTION

Massive access of IoT devices

THE IOT IS GOING TO BE BIG

THOUGH NOBODY REALLY KNOWS HOW BIG …

28.1 BILLION

Units by 2020

$1.7 TRILLION

GLOBAL SOLUTION REVENUES BY 2020

Source: May 2015

IDC
Analyze the Future

25 BILLION

Units by 2021

$200 BILLION

SERVICE REVENUES IN 2020

$1.7 TRILLION

GLOBAL ECONOMIC VALUE IN 2020

Source: November 2018

Gartner

25 BILLION

M2M connections by 2022

OF WHICH

2.6 BILLION

ARE CELLULAR

$1.2 TRILLION

GLOBAL OPPORTUNITY BY 2022

Source: January 2013

Machina Research

HOW TO HANDLE SUCH A LARGE NUMBER OF DEVICES?

A large share of IoT devices will be served by short-range radio technologies

- Unlicensed spectrum (e.g., Wi-Fi and Bluetooth)
  - Costless but …
  - Limited QoS and security requirements

A significant proportion will be enabled by wide area networks (WANs)

- Unlicensed Low Power Wide Area (LPWA): LoRa, Sigfox, …
  - Very limited demands on throughput, reliability and QoS
- Licensed spectrum: 4G, NB-IoT, 5G, …
  - Largely responsible for wireless connectivity on a global scale
  - Adapted to deliver reliable, secure and diverse IoT services.

CELLULAR NETWORK ARCHITECTURE

CONGESTION LOCALIZATION

A huge number of devices …

… but a limited number of resources (i.e., # of opportunities to connect)

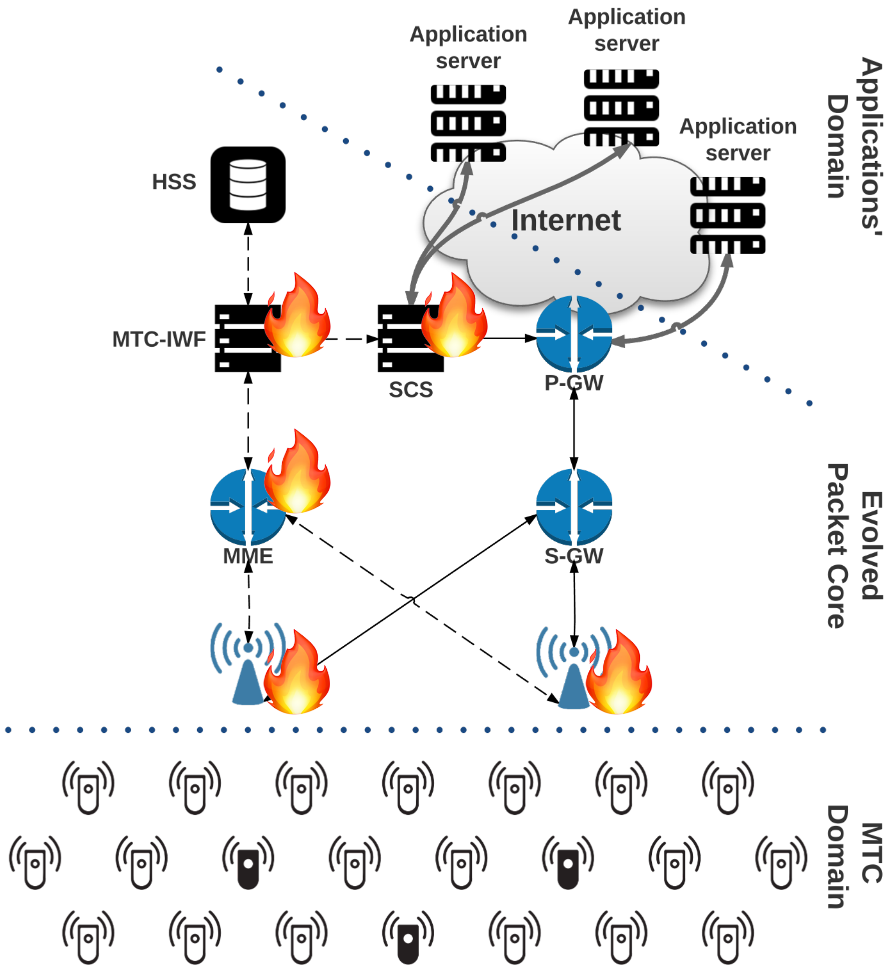
Random access

- Only way to access the network (simplest)
- **The most critical area**

Complex traffic pattern

- Poisson (e.g., credit machine in shops), Uniform (e.g., traffic lights), Beta (e.g., event driven)

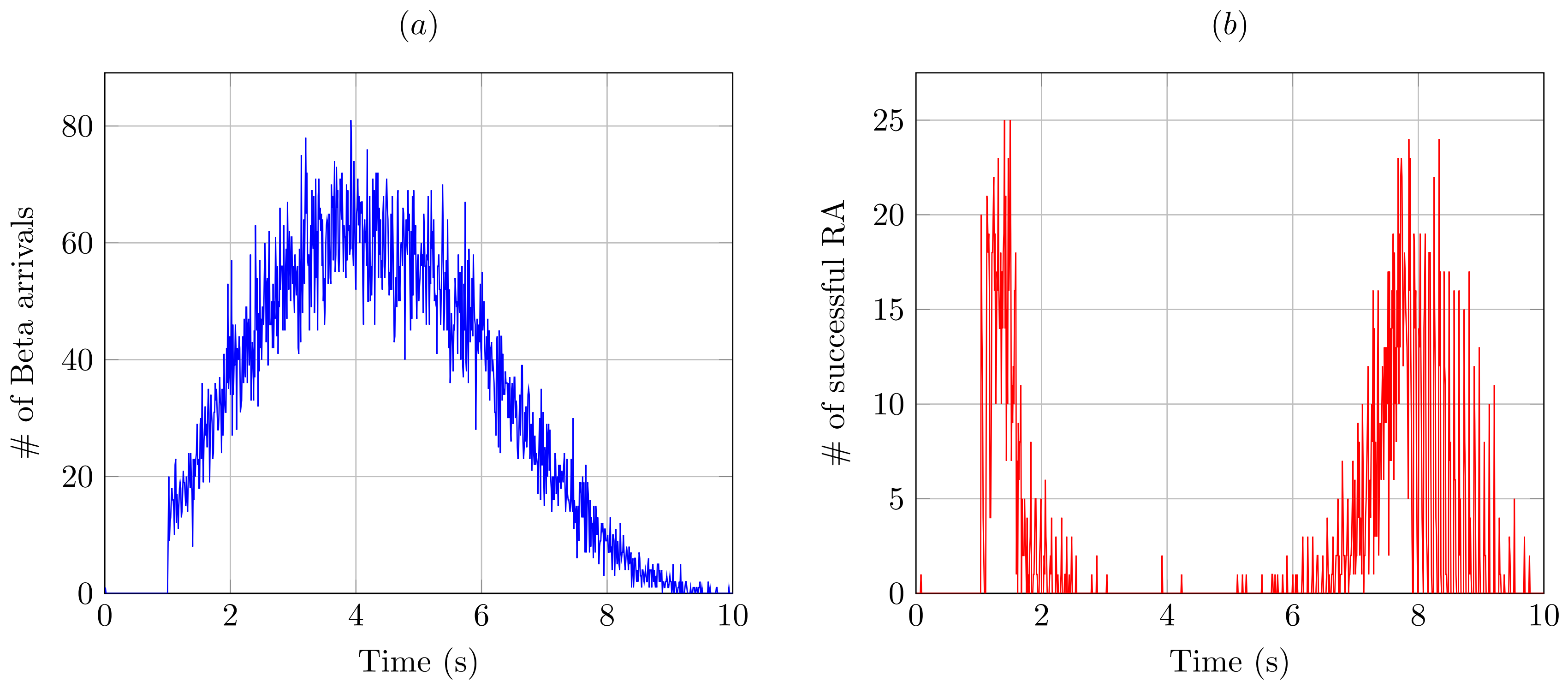
Different classes of IoT (including prioritized M2M)



Adlen Ksentini, Yassine Hadjadj-Aoul, and Tarik Taleb: "Cellular-based Machine Type Communication: Overload control". In IEEE Network, Vol. 26, Issue 6, Pages : 54 – 60 (November 2012)

RISK OF CONGESTION COLLAPSE AT THE RAN

Even when having 54 opportunities, the risk of congestion is still high …



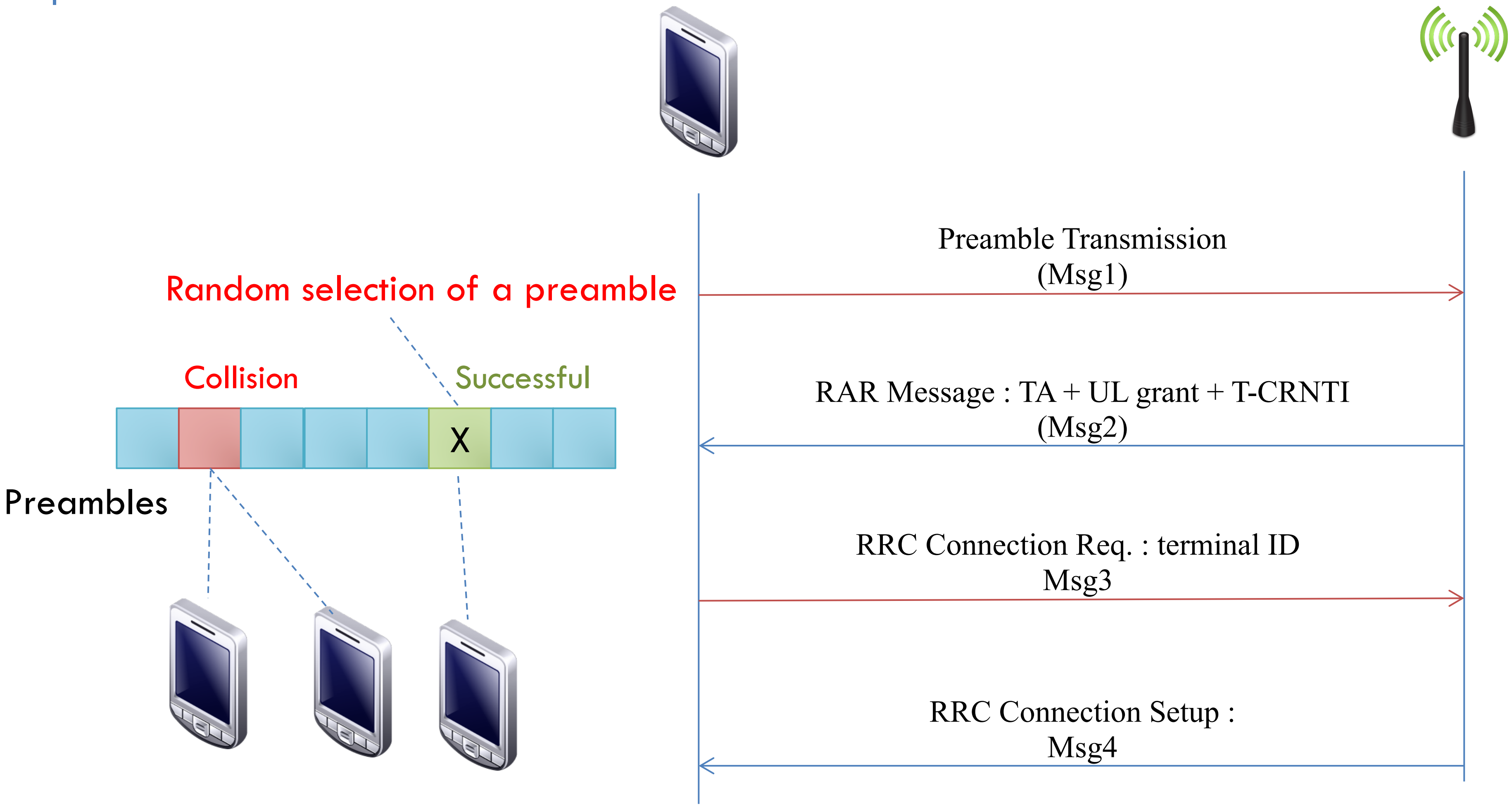
« RAN overload control … is identified as the first priority improvement area » … 3GPP TR 37.868

Meriam Bouzouita, Yassine Hadjadj-Aoul, Nawel Zangar, Sami Tabbane : "On the risk of congestion collapse in heavily congested M2M networks". In proc. of IEEE ISNCC, Hammamet, Tunis (May 2016)

ACCESS OVERVIEW OF IOT DEVICES

Understanding the origin of the problem

RANDOM ACCESS



TA: Timing Advance

T-CRNTI : Temporary Cell Radio Network Identifier

A MODEL FOR THE ACCESS

Fluid model approximating the access process

MODEL FOR ACCESS

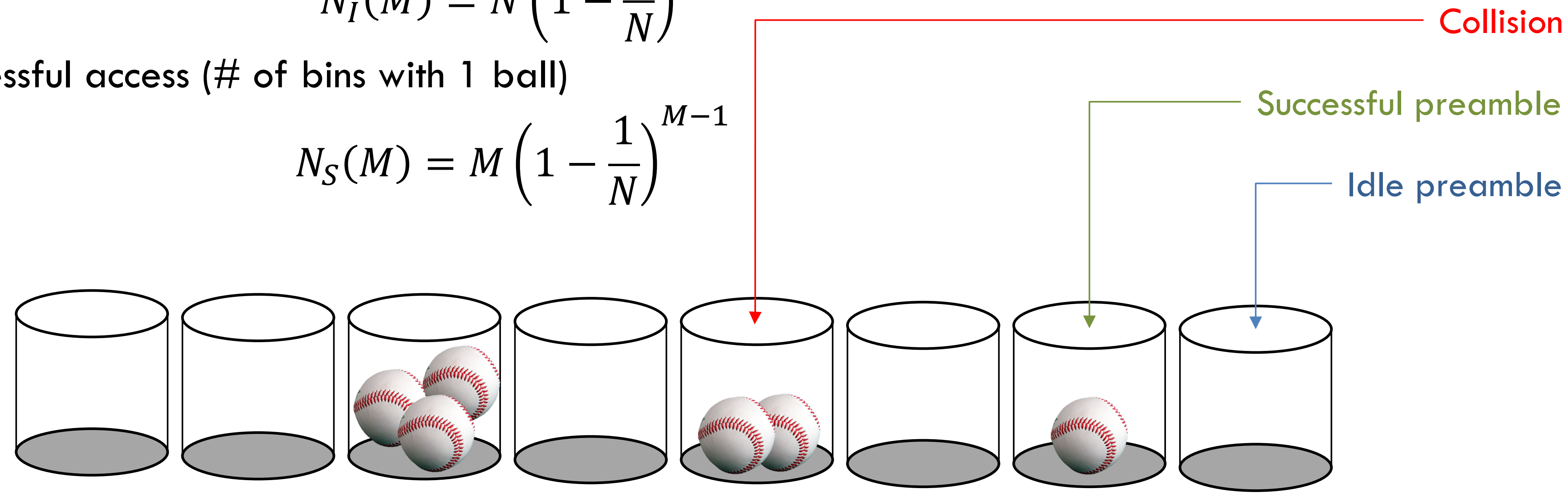
Could be modeled using the classical « Balls into Bins » problem

- N_I : # of idle preambles (# of bins with no ball)

$$N_I(M) = N\left(1 - \frac{1}{N}\right)^M$$

- N_S: # of successful access (# of bins with 1 ball)

$$N_S(M) = M\left(1 - \frac{1}{N}\right)^{M-1}$$



N Bins ~ N opportunities to connect

M Balls ~ M IoT devices

HOW TO DETERMINE THE OPTIMAL NUMBER OF CONTENDING DEVICES?

Method 1: Can be determined by Monte Carlo simulations.

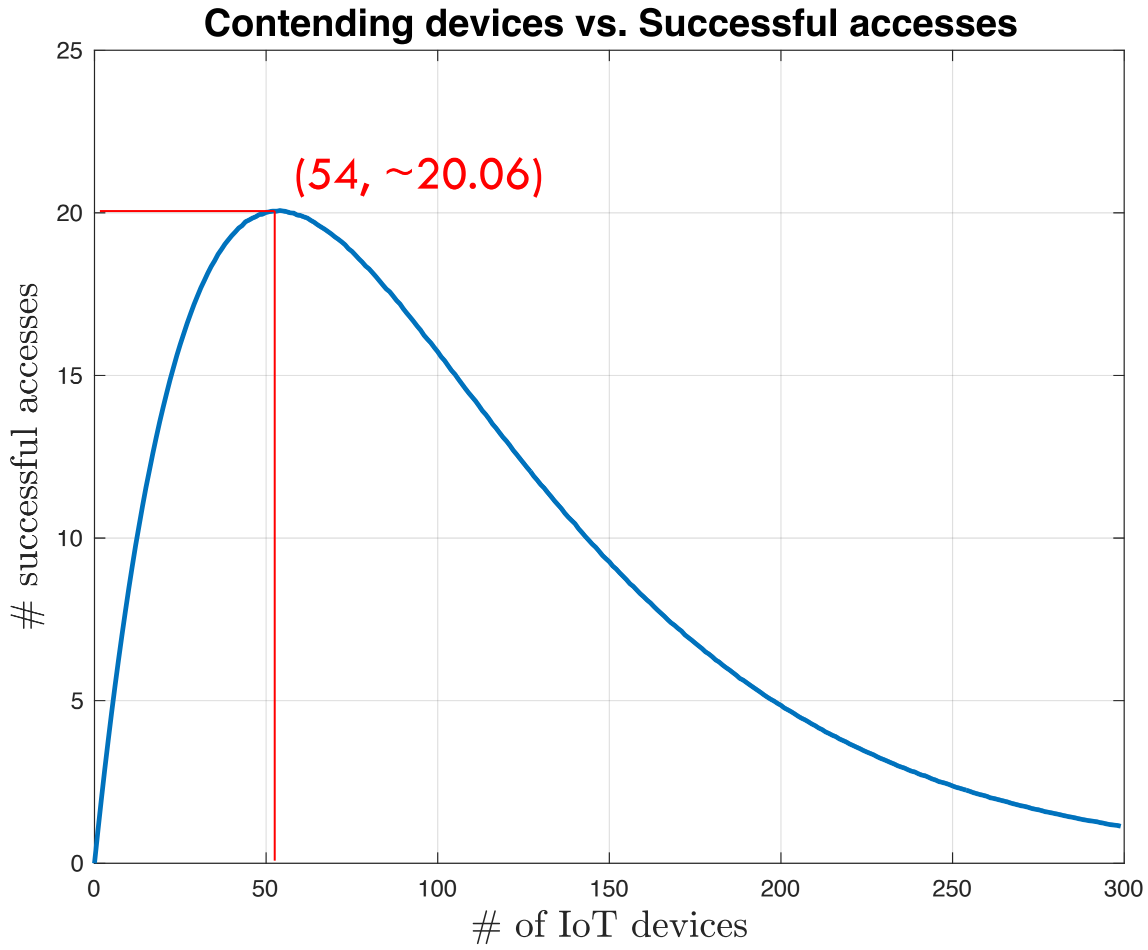
Maximized when:

# of contending devices = 54

Number of opportunities N

Method 2: Can be determined analytically.

Analysis of N_S



SOME EXISTING APPROACHES TO TACKLE THE CONGESTION AT THE ACCESS

Access planning
- Limit the burden … but insufficient since some devices react to events which cannot be timed.

Grouping devices

Pull-based scheme
- A paging message may also include a back-off time for the MTC

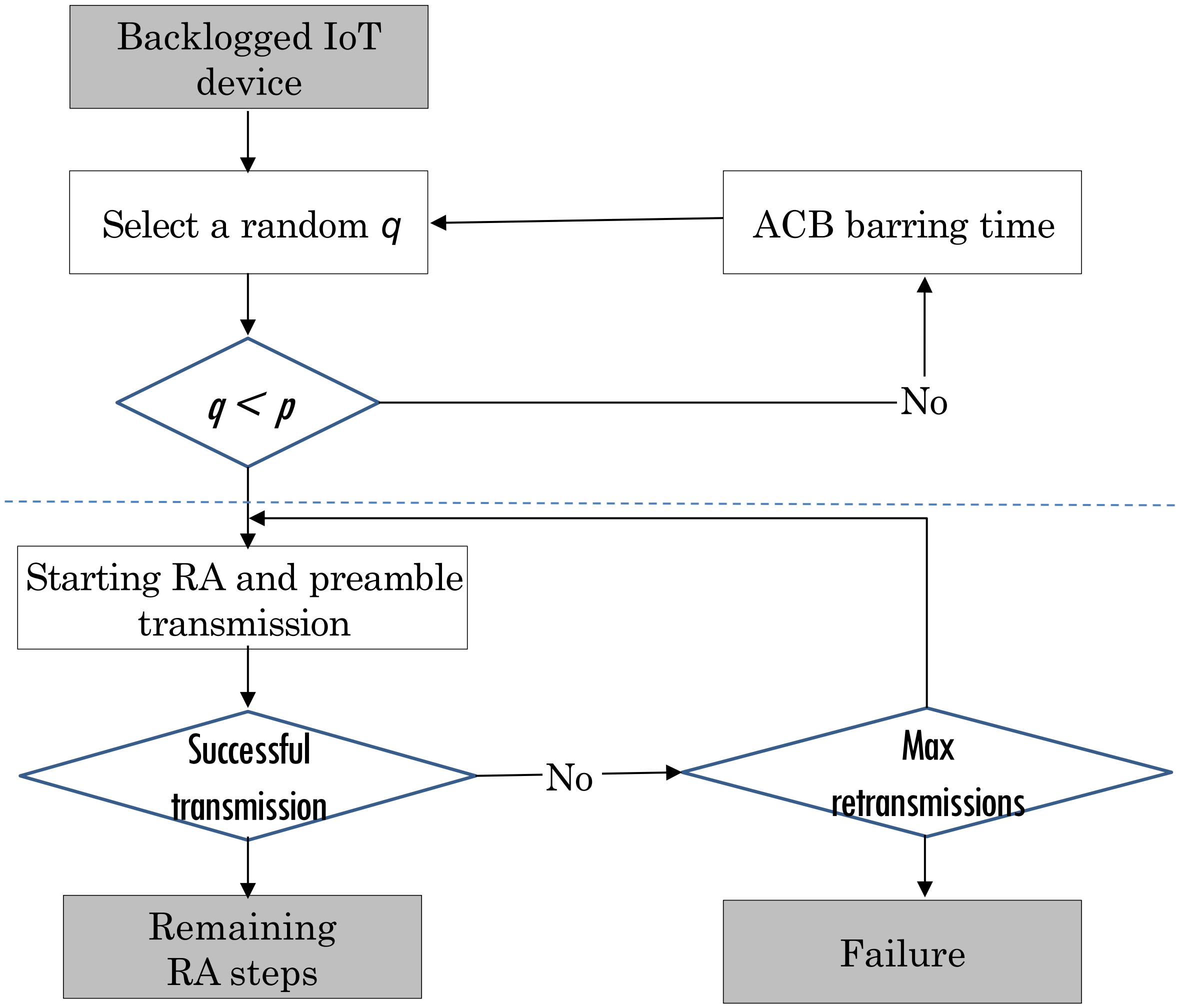
Separate RACH resources for MTC
- Splitting the preambles into H2H group(s) and MTC group(s)
- or allocating PRACH occasions in time or frequency to either H2H or MTC devices.

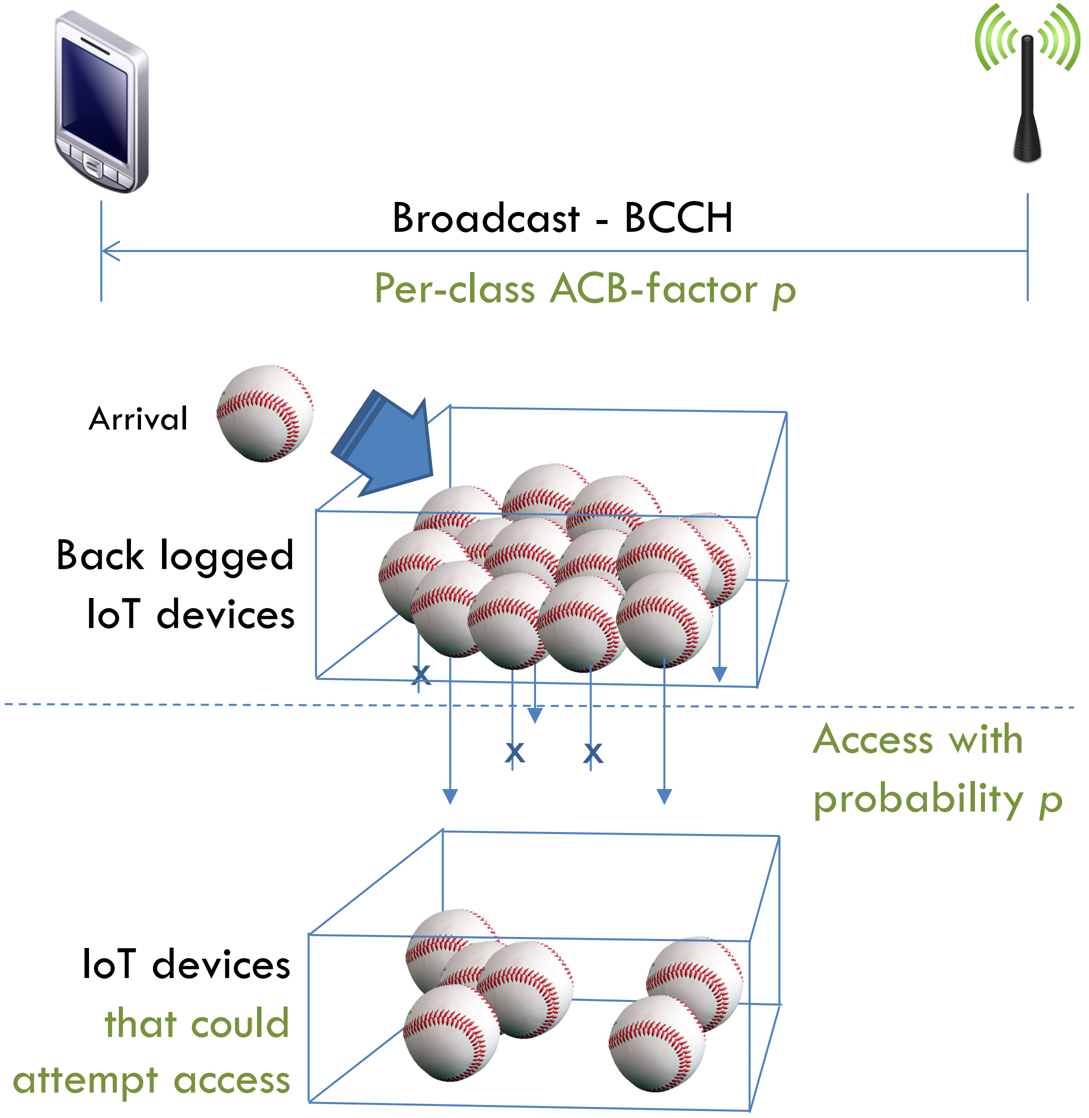
Dynamic allocation of RACH resources

Access Class Barring (ACB)
- UE individual Access Class Barring
- Extended Access Barring

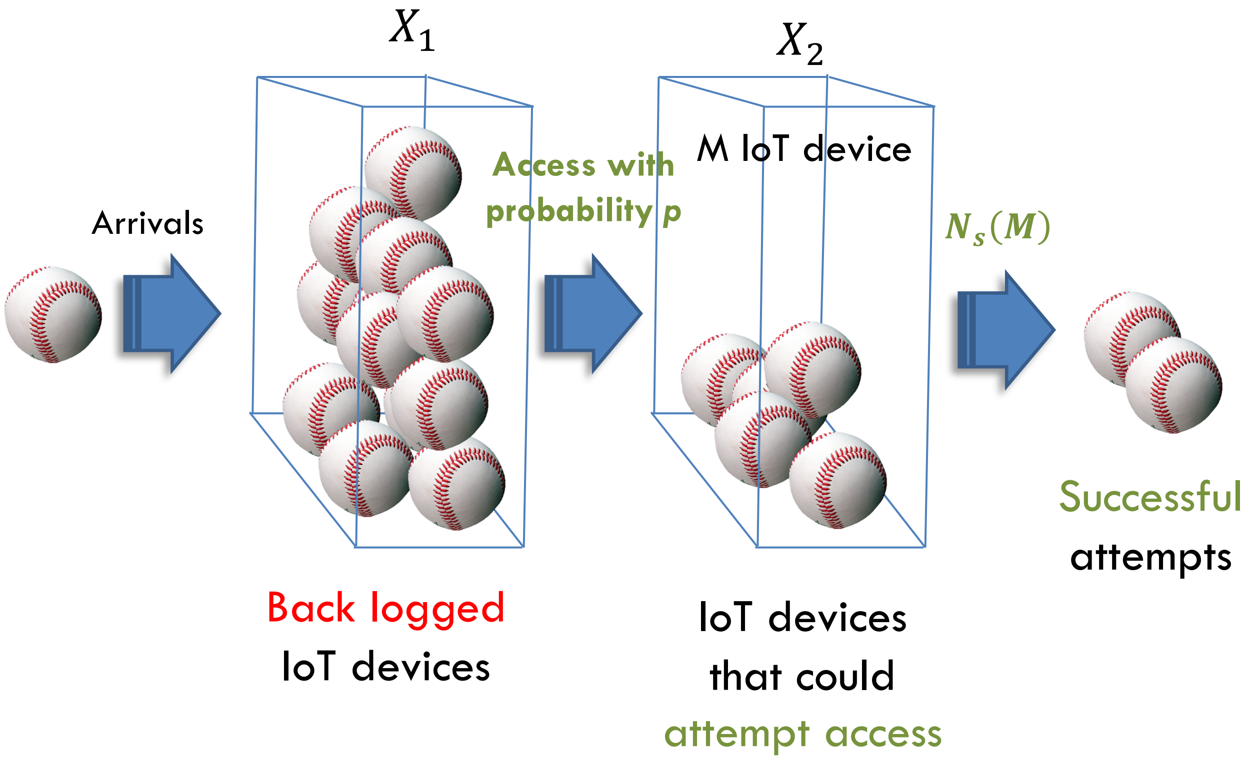
Meriam Bouzouita, Yassine Hadjadj-Aoul, Nawel Zangar, Sami Tabbane, César Viho: « A random access model for M2M communications in LTE-advanced mobile networks», In « Modeling and simulation of computer networks and systems », Elsevier/Morgan Kofmann, pp. 577 – 599 (2015)

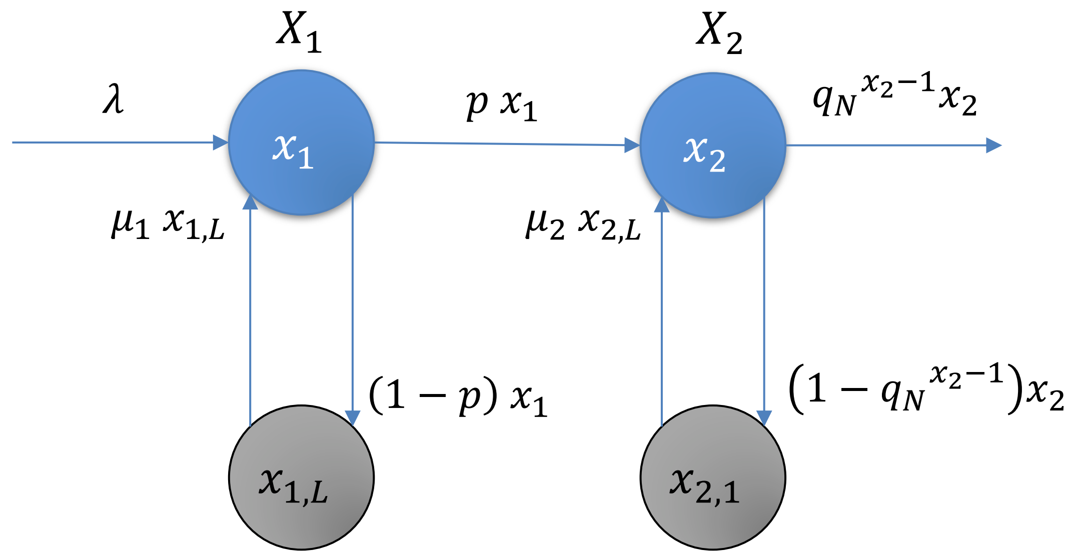
FOCUS ON THE ACB





A FLUID MODEL FOR THE ACCESS





$$
\begin{aligned}
\frac{\mathrm{d}x_1}{\mathrm{d}t} &= \lambda - x_1 + \mu_1 x_{1,L}, \\
\frac{\mathrm{d}x_2}{\mathrm{d}t} &= p x_1 + \mu_2 x_{2,L} - x_2, \\
\frac{\mathrm{d}x_{1,L}}{\mathrm{d}t} &= (1-p)x_1 - \mu_1 x_{1,L}, \\
\frac{\mathrm{d}x_{2,L}}{\mathrm{d}t} &= \left(1 - q_N^{x_2-1}\right) x_2 - \mu_2 x_{2,L}.
\end{aligned}
$$

$$q_N = \left(1 - \frac{1}{N}\right)$$

EFFICIENT SUPPORT OF IOT DEVICES

Estimating the access's contention

CHALLENGES AT THE ACCESS

What is the optimal number of contending devices

- Best target for a control strategy

How to estimate the number of contending devices (in states X_1 and X_2) ?

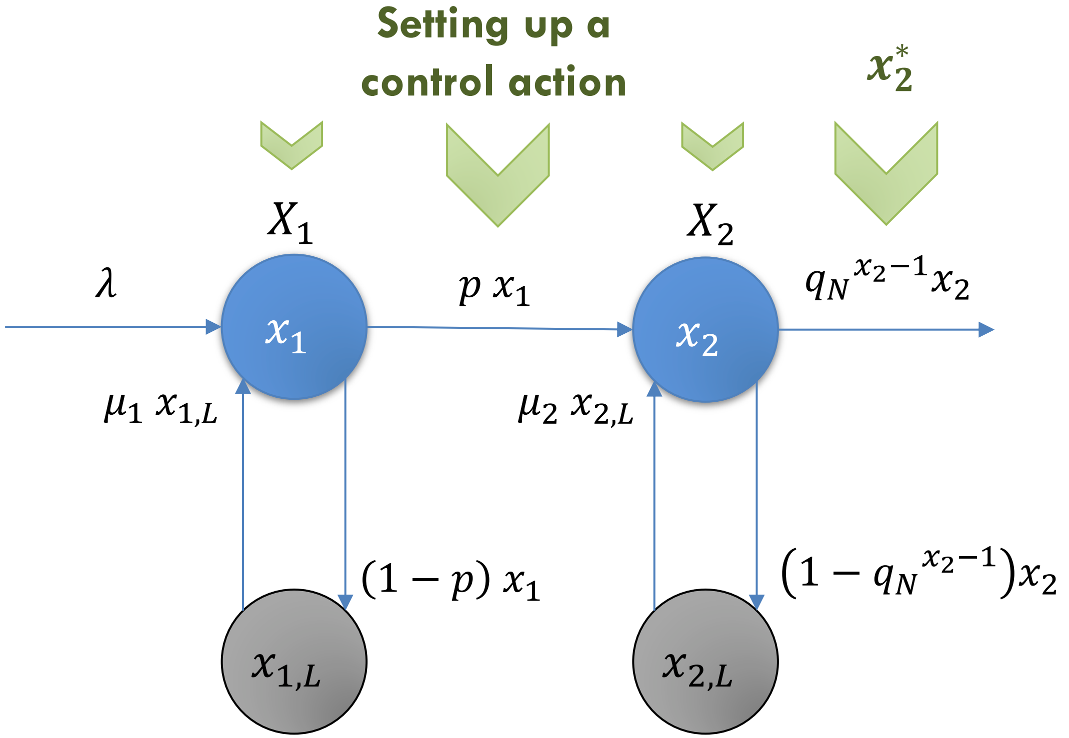
- **Difficulty:** no direct way to know it

What is the best control action to optimize the number of contending devices ?

- Optimal barring strategy
- **KPI:** delay, energy, number of abandons, number of attempts…
- **Difficulty:** Nonlinear model, non-affine in control

How prioritize the contending devices (sharing the same resources)?

- Per-class estimation, per-class barring



Meriam Bouzouita, Yassine Hadjadj-Aoul, Nawel Zangar, Gerardo Rubino: “Estimating the number of contending IoT devices in 5G networks: revealing the invisible”. In Wiley, Transactions on Emerging Telecommunications Technologies (TETT). (August 2018)

TOWARDS THE USE OF LEARNING TECHNIQUES FOR ACCESS CONTROL

WHY USING DEEP REINFORCEMENT LEARNING?

The blocking factor calculation requires a good knowledge of the number of terminals willing to attempt access

- But it is not available in the network
  - the state of the network is not observable
- It is possible to estimate this number, but this estimate is subject to noise.

The traffic pattern is very complex

Lack of data

- We cannot use supervised learning

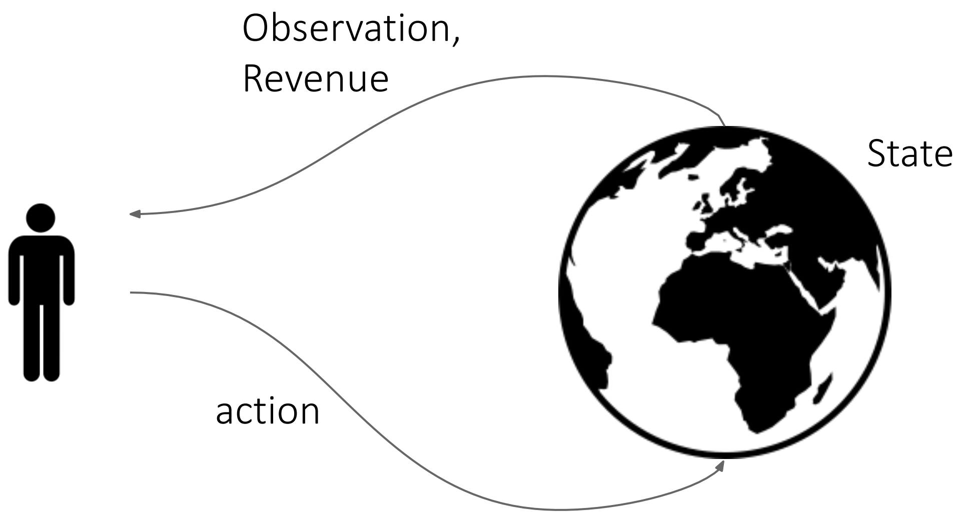
Deep reinforcement learning techniques have been shown to be effective in making predictions even when the data is very noisy.

PROBLEM FORMULATION

MARKOV DECISION PROCESS (MDP) DEFINITION

MDP: $M = (S, A, p, r)$

- State S : State space
  - $s_k = (\hat{x}_2^k, \hat{x}_2^{k+1}, \dots, \hat{x}_2^{k-H-1})$,
    - H: Horizon
    - k: time step (each new frame)
  - s_k reflects better the real state
- Action A : Action space
  - p: blocking factor
    - Continuous, deterministic
- $p(s'|s,a)$: transition probability
  - Related to the environment (not known)
- Revenue $r(s,a,s')$: is the reward of transition (s,a,s')
  - $r_k = \frac{1}{NH} \sum_{i=k-H+1}^{k} N_s^i$

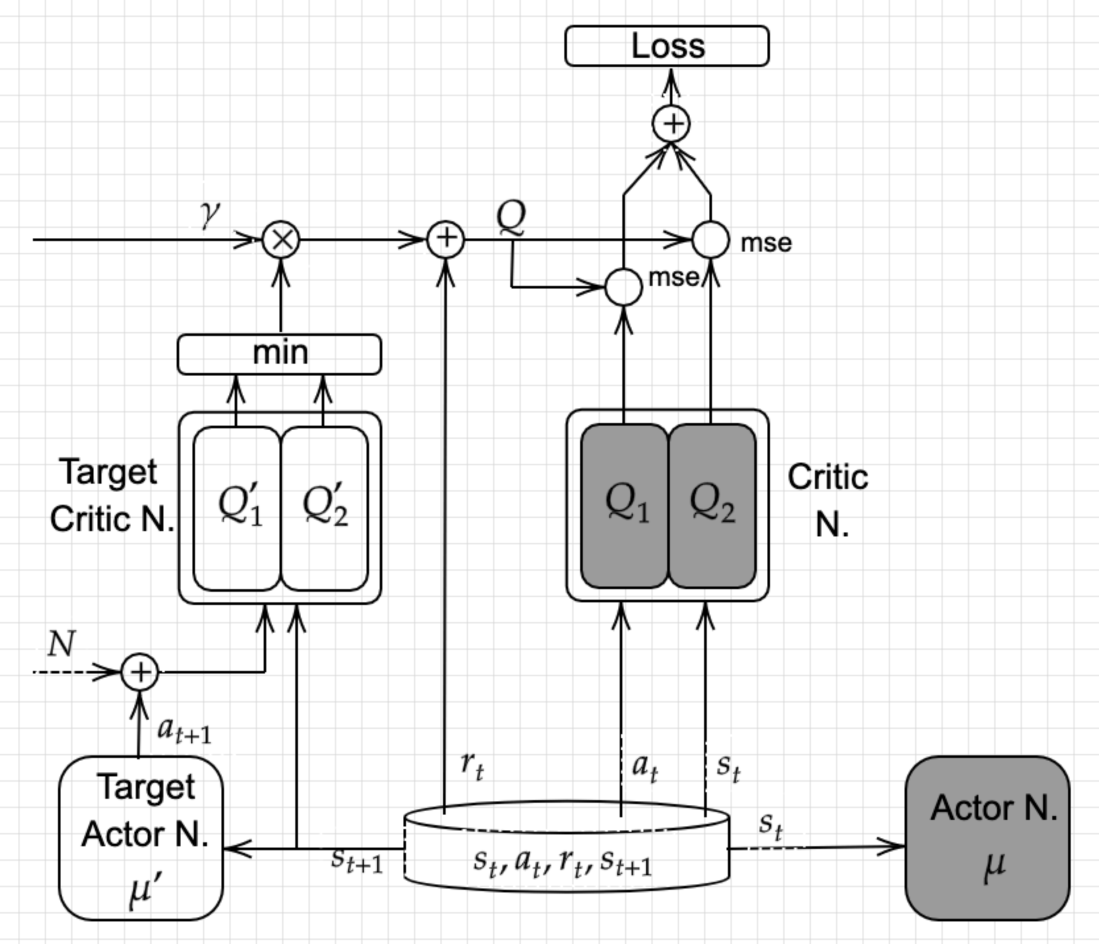


Objective:
Find the probability of blocking that maximizes the average reward.

HOW TO SOLVE THE PROBLEM?

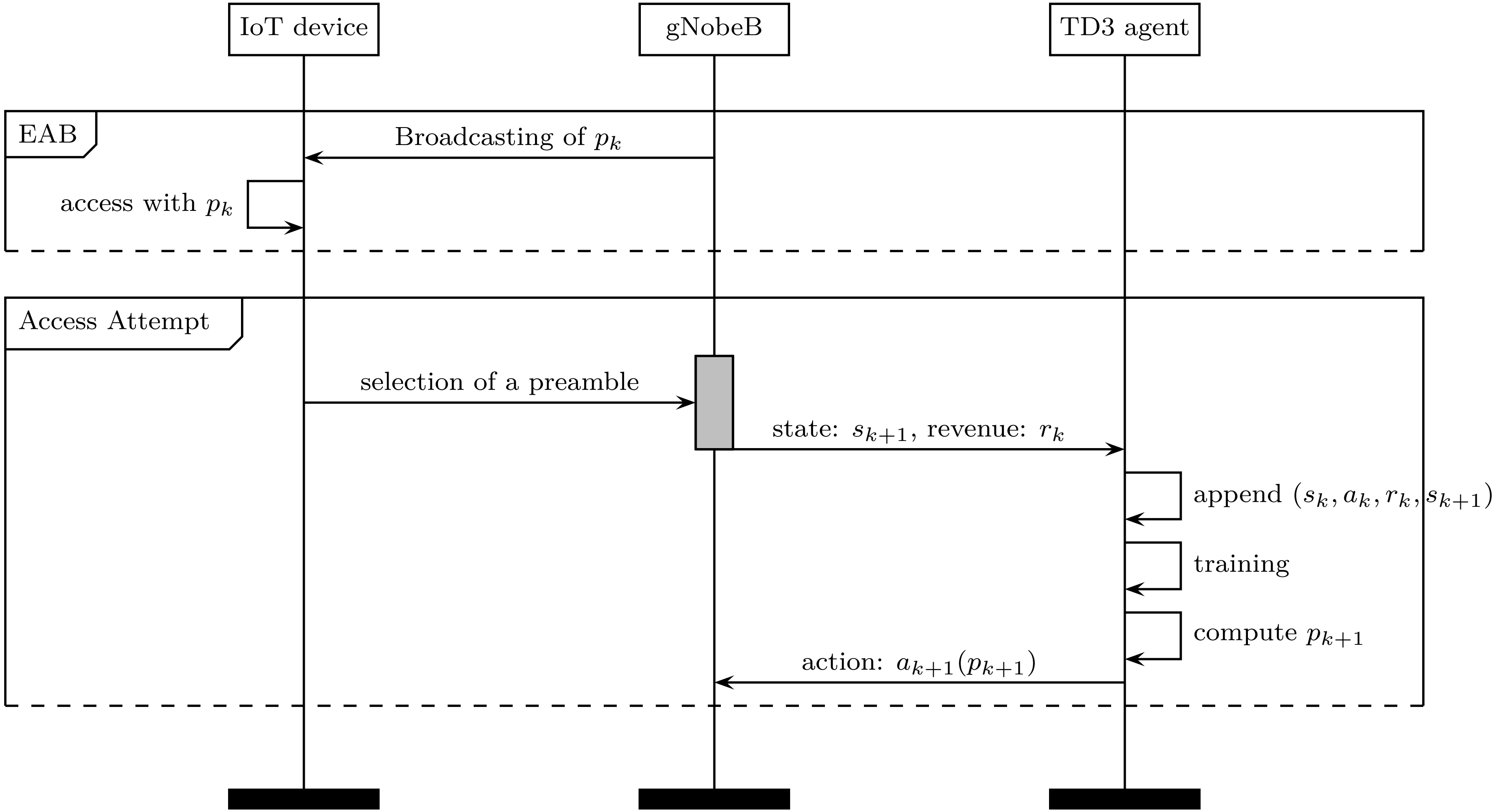
Twin Delayed Deep Deterministic policy gradient algorithm (TD3)

- Deterministic approach
- Deals with continuous action space
- Solves the problem of overvaluation in value estimation
  - Performs better than DDPG, PPO, …



Fujimoto, S.; van Hoof, H.; Meger, D. Addressing Function Approximation Error in Actor-Critic Methods. *arXiv* 2018, arXiv:1802.09477.

ARRIVAL REGULATION SYSTEM



PERFORMANCE EVALUATION

Simulator:

- Discrete event simulator developed from scratch

Arrival process of IoT devices:

- Poisson
- MTBA = 0.018s

Preambles:

- Number of preambles: $N = 12$
- Arrival frequency: 0.1s

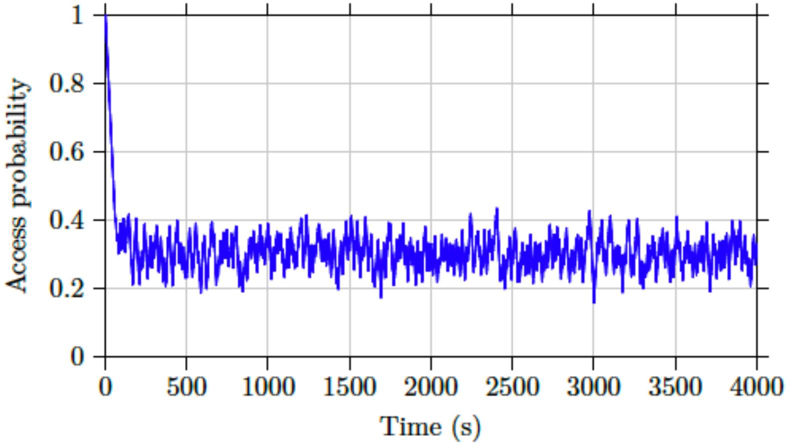
Others:

- Measurement horizon: $H = 10$

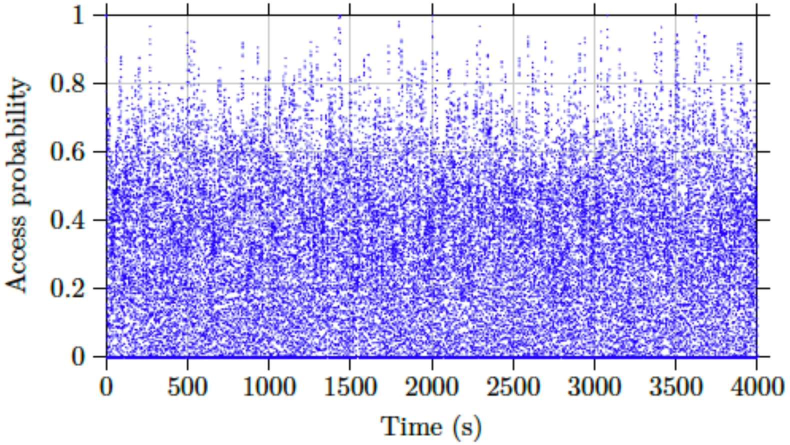
Compared strategies:

- ADAPT
- PID controller
- TD3 (proposed)

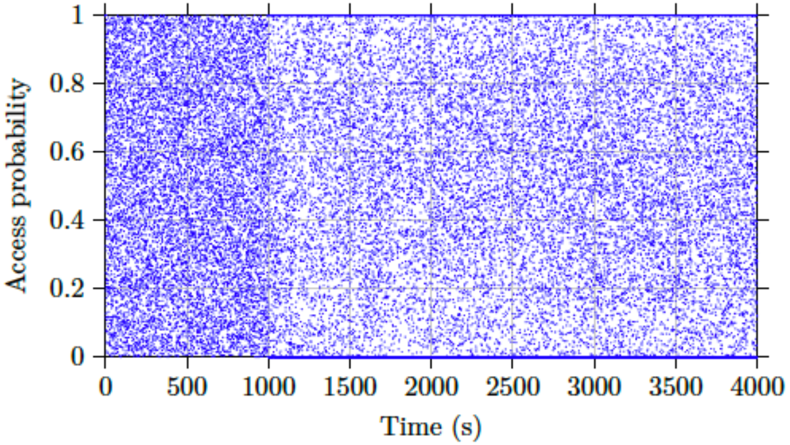
THE ACCESS PROBABILITY FOR THE CONSIDERED STRATEGIES



(a) ADAPT

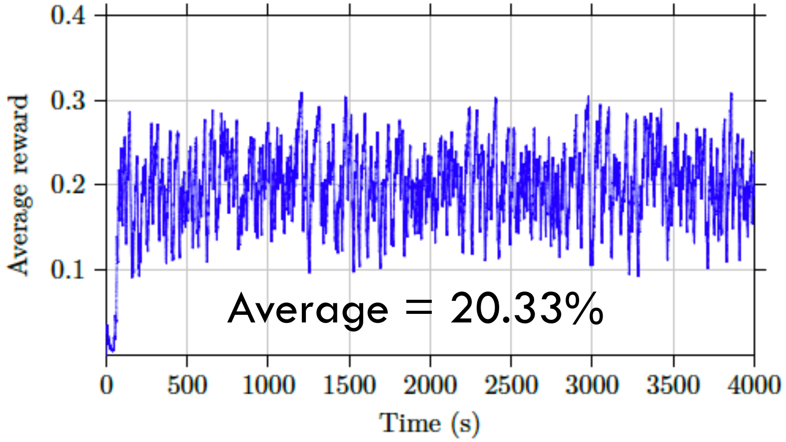


(b) PID

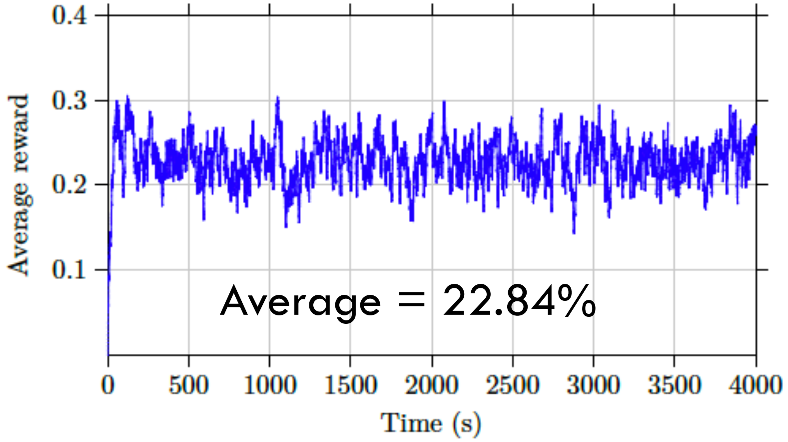


(c) TD3

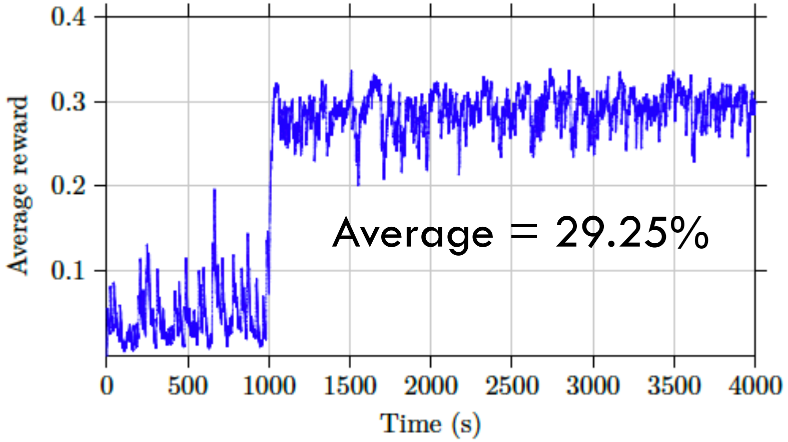
THE AVERAGE REWARD OF THE CONSIDERED STRATEGIES



(a) ADAPT



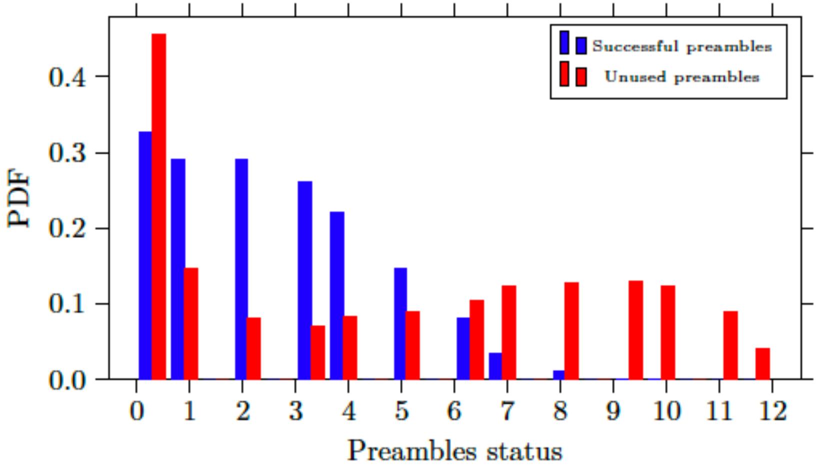
(b) PID



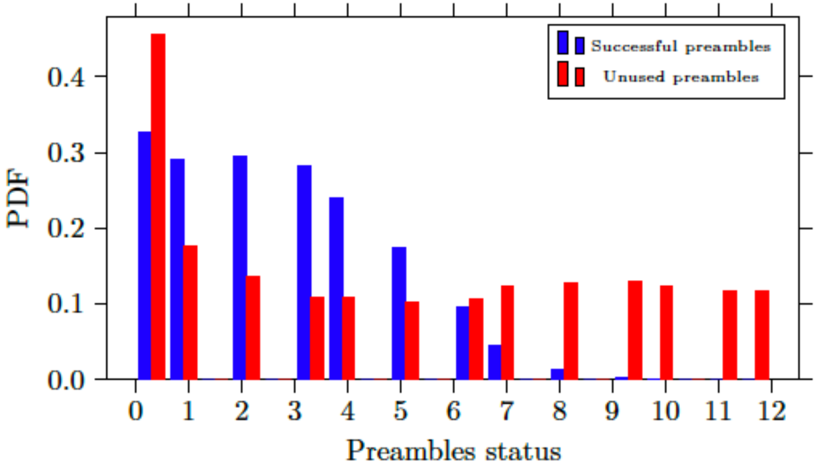
(c) TD3

THE STATUS OF THE PREAMBLES

Ave. success : 2.47
Ave. attempts: 23.52



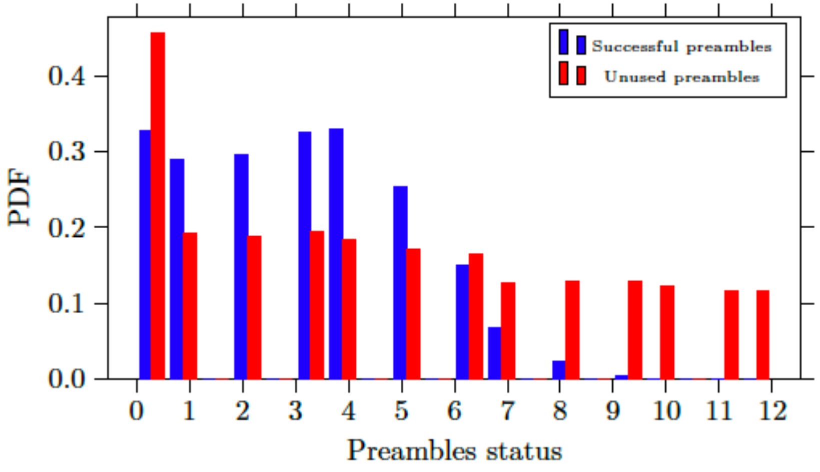
(a) ADAPT



(b) PID

Ave. success : 2.74
Ave. attempts: 17.15

Optimal success: 4.61
Optimal attempts: 11.49



(c) TD3

Ave. success : 3.52
Ave. attempts: 15.70

CONCLUSIONS

We proposed a mechanism to control the congestion of IoT access networks

- We proposed a fluid model of the access
  - Allow determining optimal objective
- We exploited recent advances in deep reinforcement learning, through the use of the TD3 algorithm

Simulation results show the superiority of the proposed approach

- Despite the lack of accurate data

Future work:

- Improve the estimation of the number of attempts.